

**THE EFFECT OF SELF-EVALUATION,
SELF-RECORDING ALONG WITH SELF-LISTENING,
AND MODELING ON
2ND GRADERS' MELODIC SINGING ACCURACY**

Dylan D. Fixmer

Submitted to the faculty of the
Jacobs School of Music in partial fulfillment
of the requirements for the degree,
Master of Music Education
Indiana University
May 2017

Accepted by the faculty of the
Indiana University Jacobs School of Music,
in partial fulfillment of the requirements for the degree
Master of Music Education

Thesis Committee

Peter Miksza, Research Director

Brent Gault

Katherine Strand

Acknowledgements

The author wishes to express his most sincere gratitude to the following people for their help, guidance, generosity, and support during this project.

Dr. Peter Miksza

Dr. Patrice Madura Ward-Steinman

Dr. Katherine Strand

Dr. Brent Gault

Sarah Maggie Olivo

Karen Bennett

Matthew VanVleet

and

My family who have supported me through my academic endeavors

Abstract

The purpose of this study was to investigate the effects of self-evaluation, self-recording along with self-listening, and modeling on second graders' melodic singing accuracy. Pre- and post-treatment singing voice data were collected from 48 second grade children from two schools in southern Indiana, who were randomly assigned to one of two treatment groups or a control group. Results indicated that children in the treatment group involving self- and model- listening improved in melodic singing accuracy whereas participants who only self-listened made no improvement and participants in the control group performed worse. Results suggested that children may benefit from the use of self- and model-listening, as well as use of self-listening as part of the self-evaluation process.

Table of Contents

| | |
|--|-----|
| Acknowledgements | iii |
| Abstract | iv |
| Table of Contents | v |
| Chapter 1 Statement of the Problem | 1 |
| Rationale | 1 |
| Problem Statement | 3 |
| Purpose Statement | 6 |
| Research Questions | 6 |
| Definition of Terms | 7 |
| Delimitations | 8 |
| Chapter 2 Review of Related Literature | 9 |
| Development of the Child Singing Voice | 9 |
| Theoretical and Practical Aspects of Musical Self-Regulation | 24 |
| Self-Perception, Self-Recognition, and Self-Awareness | 28 |
| Self-Evaluation, Rubrics, and Technology | 29 |
| Summary | 34 |
| Chapter 3 Methodology | 35 |
| Participants | 35 |
| Measures | 35 |
| Treatment | 37 |
| Procedure | 39 |

| | |
|---|----|
| Analysis..... | 40 |
| Chapter 4 Results | 42 |
| Overview..... | 42 |
| Analysis..... | 42 |
| Chapter 5 Discussion | 48 |
| Summary | 48 |
| Limitations | 50 |
| Implications for Further Research | 51 |
| Suggestions for Teachers | 53 |
| Conclusions..... | 54 |
| References..... | 56 |
| Appendices..... | 66 |
| Appendix A | 66 |
| Appendix B | 71 |
| Appendix C | 72 |
| Appendix D..... | 73 |
| Appendix E | 74 |

Chapter 1:

Statement of the Problem

Rationale

Elementary music teachers often look for ways to differentiate their classrooms and tailor their teaching to reach individual students. But, due to a lack of time and/or an overwhelming population of students, many are unable to give their students a meaningful amount of individual attention, especially when it comes to preparation of the singing voice. Self-assessment strategies can help with both individual motivation and academic achievement (Bingham, Holbrook, & Meyers, 2010, p. 59), but findings about self-assessment in music have suggested that this practice is mostly ineffective (e.g. Hewitt, 2011, p. 17). In other studies, researchers have proposed that evaluations made by peers are more accurate than self-evaluation under the same conditions (e.g. Ho, 2014). One study (Silveira & Gavin, 2016) argued that middle school children were more critical in their self-evaluations after listening to a recording of themselves, but no studies to date have investigated whether young children are able to self-evaluate when listening to recordings of their own singing.

The development of self-regulatory skills is integral to music learning (McPherson & Zimmerman, 2002). In their chapter describing the relevance of self-regulated learning theory to music learning, Zimmerman and McPherson outline the importance of self-regulation not as a fixed characteristic but “as a context specific set of processes that students draw on as they promote their own learning” (p. 328). This crucial distinction begs the questions “When does self-regulation begin?” “How is it

fostered?” and “What will aid students in their ability to self-regulate?” Some of the principles of self-regulated learning such as goal setting, motivation, self-efficacy, and self-guided practice have been addressed in recent studies (e.g. Bailey, 2006; Bandura, Caprara, Barbaranelli, Gerbino, & Pastorelli, 2003; Varela, Abrami, & Uptis, 2014). However, as far as can be determined there is no current research that deals with the explicit instruction of self-regulatory skills and habits like self-listening, as it pertains to children and the development of the singing voice.

Computer technology and mobile technology have been more prevalent in the classroom in the early 21st century and teachers in the United States have been encouraged to use these technologies in their classrooms in creative instructional ways. Although recording technology is available on most computers and practically all mobile devices, very few studies have looked at using recording technology as an unassisted aid for self-evaluating a child's singing voice. Findings from a study pertaining to the use of technology focused on how computer gaming with concurrent visual feedback affects pitch-matching accuracy indicated that such technology may improve pitch-matching ability (Paney & Kay, 2015). Since this type of technology automatically assesses the accuracy of a performance for the child, less self-regulatory process is necessary for the participant to make progress. The effectiveness of basic recording technology in helping children to develop self-regulatory ability and melodic singing accuracy skills by self-listening is therefore the basis of this study.

Problem Statement

Many studies have been conducted on how children develop their singing voices (Goetze, Cooper, & Brown, 1990) especially in relation to the use of modeling as a vocal aid, and the use of unison group singing versus individual singing. Modeling has been found to be helpful for vocal development, with child models and female models with minimal vibrato being the most effective (Goetze, Cooper, & Brown, 1990; Green, 1990; Hedden & Baker, 2010; Rutkowski & Miller, 2003; Welch, Sergeant, & White, 1997; Yarbrough, Green, Benson, & Bowers, 1991). This evidence suggests that a prerecorded model (both child's voice and woman's voice) made available for participants to listen to could affect how the participants engage in self-regulatory behaviors.

Studies of unison group singing suggest that children may also benefit from individual singing (Green, 1994; Rutkowski & Miller, 2003). Although Green (1994) concluded that children sing more accurately together in small groups of eight, she discussed the need to further study how children sing alone. Studying how children individually sing when given the opportunity to reflect on prerecorded models and recordings of themselves may therefore be beneficial.

Researchers examining pitch-matching accuracy have done so in order to study children's singing development (e.g. Dalla Bella, 2015; Demorest & Pfordresher, 2015). However, Mang (2006) discussed important distinctions between pitch-matching and melodic singing accuracy, stating:

pitch-matching as the criterion task might not be readily generalised to children's singing performance of songs. It should also be noted that singing, not pitch-

matching, is one of the central activities of the elementary school music curriculum, and that song performance can provide a much more musical expressive experience for children (p. 163).

For this reason, studying children's ability to sing melodies accurately may be more beneficial than simply pitch matching ability.

Practitioners in music education and researchers outside of music writing on self-reflection and self-evaluation in the elementary classroom have provided conflicting recommendations (e.g., Bingham, Holbrook, & Meyers, 2010; Ross, 2006; Wells, 1998), but all have suggested that a student's ability to self-evaluate can be a helpful part of music learning. Uses of self-evaluation strategies have also been studied and have suggested the use of either teacher- or student-generated rubrics (e.g., Eppink, 2002; Wesolowski, 2012; Whitcomb, 1999). Ross (2006) noted that self-evaluation techniques work best when students are trained how to assess their work using tools such as rubrics. A significant relationship has been found between the use of self-reflection and the development of musical abilities especially when a model is provided (Davis, 1981; Hewitt, 2000). These authors along with several others (Ho, 2004; Kenny, 1998; Kostka, 1997; Wells, 1998) have stressed the importance of self-regulated practice as an integral aspect of successful musicianship.

Research of self-regulated music learning suggests that students may begin to develop self-regulatory behavior around second grade. McPherson and Zimmerman in *The New Handbook of Research on Music Teaching and Learning* (2002) indicated that self-regulation begins to develop at the second-grade level, which could suggest that

eight to nine year-old children would be an appropriate range to test assisted self-regulatory practices. They concluded that more research on how students specifically develop self-regulatory abilities would greatly benefit the music education profession.

There is a possibility that children will be unable to recognize their own voice or may have some degree of congenital amusia which would cause them to be physically unable to use the technology in this way. However, according to research the probability that a child will have congenital amusia or other musical deficiencies is lower than 3% (Cuddy, Balkwill, Peretz, & Holden, 2003) and children as young as four years tend to be able to recognize their own voices (Strömbergsson, 2012). The likelihood of children not being able to benefit from self-evaluation due to amusia or an inability to recognize their own voice is small.

Computer and mobile technology have been successfully used in classrooms to aid in self-regulated learning. For example, there are instructional tools available to help students practice and identify errors in their performances (e.g., SmartMusic). Pitch-matching assessment games have also shown to improve students' vocal accuracy with prolonged use (Paney & Kay, 2015), and Silveira and Gavin, (2016) found that using recording for self-listening had an effect on middle schoolers' pitch accuracy in self-assessment. However, as far as can be determined, there are no studies that have investigated how recording technology could help develop self-regulation, self-evaluation, and vocal accuracy skills in children.

Varela, Abrami, and Upitis (2014) suggested that based on current understandings of self-regulation research the process of self-recording using technology may provide

retroactive mastery evidence and bolster self-efficacy beliefs. Based on the previous literature it stands to reason that the development of these three elements through the use of computer or mobile technology recording should be examined. Furthermore, the Music Model Cornerstone Assessments developed by the National Association for Music Education in 2015 includes a sequence of assessments for second-grade students for creating and performing music that includes self-recording, self-listening, self-assessment, and refinement. These documents show that the music profession values student self-evaluation specifically through the process of recording as a form of assessment measurement of student achievement. It is important that such values be supported by research pertaining to children's ability to self-regulate in order to validate or invalidate these assessments.

Purpose Statement

The purpose of this study was to investigate the effects of self-evaluation, self-recording along with self-listening, and modeling on second-graders' melodic singing accuracy.

Research Questions

- What are the effects of the following instructional conditions on melodic singing accuracy:
 - Use of only a rubric for self-evaluation?
 - Use of a rubric, and self-listening?
 - Use of a rubric, self-listening, and listening to a recorded model?
- Is there a significant difference in achievement between each of the above conditions?

Definition of Terms

Melodic Singing Accuracy: The ability to maintain tonality (i.e. sing in tune) during a performance of a song (Mang, 2006, p. 162). Melodic singing accuracy differs from pitch-matching ability in that it does not ask for the participant to echo individual pitches from a auditory stimulus, but instead asks him/her to recall tonality from memory.

Modeling: “listening to an "ideal" audiotaped version of the music being performed for the purpose of emulating the exemplar performance” (Hewitt, 2000, p. 17).

Rubrics: “A set of scoring criteria used to determine the value of a student's performance on assigned tasks; the criteria are written so students are able to learn what must be done to improve their performances in the future” (Asmus, 1999, p. 21).

Self-Evaluation: How students “respond to feedback, monitor their own progression, and evaluate how effectively they are learning” (McPherson & Zimmerman, 2002, p. 337).

Self-listening: Listening attentively to one’s own performance (Hewitt, 2000, p. 17) for the purpose of self-evaluation.

Self-Regulated Learning (SRL): “Self-observation to strategically adjust one’s performance process or method of learning” (McPherson & Zimmerman, 2002, p. 328).

Delimitations

The results of the current study will be generalizable to second-grade students in public elementary school music classrooms in southern Indiana. Children in second grade have been selected for participation as it has been suggested that children begin to show signs of self-regulatory skills around age eight (McPherson & Zimmerman, 2002).

Chapter 2:

Review of Related Literature

The following review of related literature covers studies pertaining to the development of children's singing voices, theoretical and practical aspects of musical self-regulation, the ability for children to hear and recognize their own voices, and how self-evaluation and technology are being currently used to develop musical abilities.

Development of the Child Singing Voice

The following studies and reviews collectively illustrate our understanding of how children's voices develop and how pitch-matching and melodic singing-accuracy data are collected. Major factors affecting pitch accuracy include quality modeling, reliable assessment, unison vs. group singing, frequency of instruction, use of singing voice, age, gender, and tonal aptitude.

Goetze, Cooper, and Brown (1990) reviewed research dealing with children's singing ability, factors influencing poor singing ability, and processes required for accurate singing (i.e., pitch discrimination, pitch production, pitch monitoring, and motivation). Their synthesis revealed that presence of models, singing task characteristics, text characteristics, individual vs. group singing settings, and use of accompaniment were found to be important factors pertaining to singing accuracy. Most relevant to the current study, Goetze et al. (1990) concluded that the presence of a model can "inhibit or enhance a child's ability to sing accurately" (p. 23) based on the type of voice (e.g. female adult) and quality of the voice, and that asking a child to model for the class may be the most beneficial (p. 31). Experimenting with pedagogical methods that

involve a modeling component could be particularly beneficial, since the juxtaposition of modeling and self-listening has not yet been studied. Goetze et al. (1990) also identified several problems with the extant research in children's singing and suggested that methodological inconsistencies related to measurement, (i.e. rating scales or oscilloscopic devices) were particularly problematic because each method varies in reliability and precision. Advances in digital pitch analysis software since 1990 can allow for measurements to be much more exact and reliable.

To examine some of these aspects of singing ability and acquisition further, studies pertaining to developing accuracy indicate that singing accuracy can be affected by the number of children singing, sex and age of the children, amount of instruction and experience, and type of singing voice used.

Green (1994) investigated differences in vocal pitch accuracy as a function of unison vs. group singing among elementary aged children. In her rationale, Green (1994) noted conflicting results from previous studies. Evidence suggesting that students match pitch more accurately when singing alone (e.g., Bently, 1969) and studies suggesting unison singing with others (e.g. Gould, 1969) is more effective created cause for further investigation due to the importance placed on singing in the general music classroom.

Participants were 241 children in grades 1, 2, 3, and 5 from one inner-city elementary school. Prior to collecting data, Green (1994) visited the school to help acquaint the students with the singing material and data collection process. The students were taught a simple pentatonic song in D by rote. Students were given an opportunity to review the song at the beginning of each data collection session. Pitch accuracy data

were then collected from individual students and students in groups of four via a recording device. Green (1994) evaluated each participant's recording based on accuracy of the 17 pitches in the song as well as the 16 intervals, which resulted in a possible total pitch accuracy score of 33. Each participant received one pitch accuracy score for singing the song in a unison condition and another score for singing alone.

The data were analyzed using a three-way analysis of variance with repeated measures of grade by gender by performance condition. Females (16.89) had significantly higher mean scores than males (13.22) ($p < .0004$). The mean scores for each grade level were significantly different ($p < .0008$), and tended to increase (first grade 12.54, second grade 14.51, third grade 15.36, fifth grade 18.33). Group singing was also significant ($p < .0001$) and had a higher mean score (16.42) than individual singing (13.83). Green (1990) also noted that the participants seemed much more comfortable singing together in a group than singing alone. These findings help support investigating whether or not singing with recording technology can be used as an aid for children to develop their voices.

In another study pertaining to group singing, Rutkowski and Miller (2003a) examined the effectiveness of small group singing activities and frequency of instruction on musical aptitude and singing accuracy among first graders. The two questions posed by Rutkowski and Miller (2003a) were:

1. Will frequency of instruction and/or large-group versus individual/small-group instruction affect music aptitude scores, as measured by the Intermediate

Measures of Music Audiation (IMMA) (Gordon, 1986) halfway through the treatment period and at the end of the treatment period?

2. Will frequency of instruction and/or large-group versus individual/small-group instruction affect the children's use of singing voice, as measured by the Singing Voice Development Measure (Rutkowski, 1984, 1986, 1990b, 1996, 1998) at the end of the treatment period? (p. 25).

Participants were 94 first graders from four intact classes from a single elementary school. Classes 1 and 2 received music instruction from their music teacher for 40 minutes each week while classes 3 and 4 received the music instruction for two, 20-minute periods every week. Classes 1 and 3 participated in large group activities while classes 2 and 4 participated in small group or individual singing activities. The IMMA test was administered prior to, halfway through, and at the end of the treatment to evaluate the participants' tonal aptitude. The Singing Voice Development Measure test was administered prior to and at the end of the treatment to evaluate the children's singing abilities. The participants' singing was recorded and analyzed by the teacher and another judge. Sixteen performances were repeated on each judge's tape to allow for an assessment of inter-judge reliability.

The researchers found statistically significant ($p \leq .05$) differences on IMMA scores at mid-test as a function of frequency of instruction, the mode of instruction, and the interaction of these factors. The results proved puzzling to Rutkowski and Miller (2003a) who attributed the significance at the mid-test to outside factors. Ultimately, the authors concluded that the frequency of instruction did not significantly affect singing

accuracy in small group or individual singing activities. Frequency of instruction will therefore not be included in the current study as an influential factor.

Welch, Sergeant, and White (1997) examined how gender, age, and a singing task affected pitch-matching among children. Welch, et al. (1997) stated that studies of “out of tune” singing were prevalent in the latter half of the 20th century with common findings such as females have better pitch accuracy than males and children’s voices improve with age are common (p. 153). The authors asserted that these studies were a “snap-shot” (p. 154) and do not provide longitudinal data to support their findings.

To study the longitudinal affects, Welch, et al. (1997) conducted a three-and-a-half-year study with participants ranging from four to eight years old. The researchers used two different singing assessment tasks (songs, and individual pitch patterns and fragments) which were recorded for later evaluation by six trained judges. The judges’ ratings of participants’ vocal accuracy indicated that the mean score for boys linearly declined over the three years and girls’ mean scores did not change. The reason for girls’ greater pitch accuracy over time was unclear, but could have been due to modeling and vocal identification of students to teachers (Welch, et. al., 1997). The authors concluded that participants from the study, both male and female, entered schooling with comparable vocal accuracy which was maintained through the first three years of schooling, and changes started at age seven. Based on these conclusions it may be pertinent to the current study to observe how children of each gender approach and execute the singing task.

In a study of another important facet of singing accuracy development, Hornbach and Taggart (2005) examined the relationship between developmental tonal aptitude and singing achievement among kindergarten through third-grade students. The authors' rationale stemmed from the "recent decline in in the number of elementary children who are able to use their singing voices." (p. 323). The authors identified numerous variables that may affect singing achievement including age, gender, vocal model, individual vs. group singing, and the use of accompaniment, as well as the conflicting evidence surrounding the effective relationship between musical aptitude and singing achievement.

Hornbach and Taggart (2005) collected data on 162 students from two public elementary schools using the Primary Measures of Music Audition for aptitudes and a five-point rating scale for singing accuracy. The participants performed a teacher-taught piece which was recorded and evaluated by three judges. Data were collected based on grade level and by school. The inter-judge reliabilities ranged from .76 to .97. The data for each grade level indicated that each subsequent grade level increased in mean score with the exception of third grade which scored lower than second grade in both schools. The mean scores in School 2 were higher than School 1 in all grade levels. None of the correlation coefficients relating musical aptitude and singing accuracy were statistically significant regardless of the grade level.

The authors concluded that "singing achievement does not seem to have a meaningful relationship to tonal music aptitude, regardless of age or school setting" (Hornbach & Taggart, 2005, p. 328); however, the findings do suggest that singing instruction may have a positive effect on singing accuracy. Hornbach and

Taggart (2005) also noted that the decline in third-grade singing ability may be attributed to social pressures, especially with boys, which is consistent with Welch, et. al. (1997).

The conclusions from this study suggest that musical aptitude is not an indicator of children's vocal ability and therefore some other facet of voice development should be studied.

Rutkowski (2015) studied one such facet by examining how the use of vocal registers affects singing accuracy among early elementary school children. Rutkowski (2015) hypothesized that there may be a relationship between pitch accuracy and pitch discrimination as students age into adolescence because previous research (e.g. Phillips & Aitchison, 1997) suggested that children who have difficulty matching pitch but do not have difficulty discriminating pitch may lack physical vocal coordination. Rutkowski (2015) also cited Trollinger (2003) who found a positive correlation between higher speaking voice and greater pitch accuracy, as well as between a wider speaking range and a wider singing range. Rutkowski (2015) asserted that "children may not sing accurately just because they are not yet physically comfortable with a particular register" (p. 285). Before this study, the relationship between children's vocal registers and pitch accuracy had not been empirically investigated.

Rutkowski (2015) posed three research questions:

1. What is the relationship between kindergarten and first-grade children's use of singing voice and singing accuracy?
2. Do kindergarten and first-grade children sing patterns more accurately if the pitches of those patterns fall within their accessible registers?

3. What is the relationship between first-grade children's tonal aptitude scores and their singing accuracy or use of singing voice? (pp. 285-286)

Using recordings from previous studies, 38 first grade and 37 kindergarten singing voice usages were assessed by two judges using the Singing Voice Development Measure (SVDM). Participants individually echoed an adult female singing voice in three-tone patterns on both a neutral syllable and text to eliminate conditional bias. The same recordings were also rated for pitch accuracy in terms of correct number of pitches by two different judges. Tonal aptitude scores of the first-grade sample were also taken using the IMMA.

Inter-judge reliability scores ranged from .85 to .97. When scores for combined grade levels using text versus a neutral syllable were analyzed, the difference was found to be significant ($p < .001$). The data were therefore collapsed across grades but text and syllables kept separate. Singing accuracy and SVDM scores were found to have a strong correlation ($r = .82$ for text and $r = .83$ for neutral syllable).

Rutkowski (2015) divided the data to examine how the SVDM score would affect pitch accuracy based on the child's associated register:

Group 1 was classified as singers (SVDM > 4.5 , $n = 39$) and could sing all eight patterns, Group 2, the initial range and inconsistent singers (SVDM $> 3.5-4.5$, $n = 1/4 46$), had six of the eight patterns in their register and Group 3, limited range and inconsistent initial range singers (SVDM $> 2.5-3.5$, $n = 1/4 31$), only had two patterns in their usable range. (p. 288)

Two MANOVA calculations were used to examine the difference between singing accuracy within and outside of available register. Both analyses indicated that children were significantly more accurate when singing within their available register based on their SVDM score. The data supported Rutkowski's (2015) hypothesis that there were statistically significant positive correlations between usable singing register and singing accuracy. Rutkowski (2015) suggested that a child's use of singing voice should be carefully considered when evaluating singing accuracy and that helping children access their full vocal registers may lead to more accurate singing.

The previously cited studies of pitch accuracy development have focused on singing accuracy, but only of a particular age group. Demorest and Pfordresher (2015) were interested in collecting data from participants ranging from kindergarten through adulthood using a similar set of tasks and identical scoring procedures. They examined the "differences in singing accuracy between children of different ages and adults on matching and song singing tasks," and "the relationship between scores derived from human judgments and those generated by an acoustic analysis" (p. 295). Data from three previous studies (Demorest, Nichols, & Pfordresher, 2014; Demorest, Pfordresher, & Kelley, 2014; Pfordresher & Brown, 2007) were used to assess singing accuracy data on 78 adults, 55 sixth graders, and 77 kindergarteners. All three investigations assessed pitch accuracy by single pitch, interval, and pattern pitch echoing. The data collected also included recorded scores for participants singing a memorized familiar song. Each participant's recorded voice was evaluated using an eight-point scale developed by Wise and Sloboda (2008) and via an acoustic analysis of pitch.

The results from the investigation showed significant ($p < .001$) differences between the sixth-grade participants and the other two groups, but no difference between adults and kindergartners. Task complexity was found to also be a statistically significant ($p < .001$) factor with single pitch being the easiest, followed by interval pitch, and then pattern. The results suggested that age does not affect pitch accuracy. The authors suggest that this puzzling conclusion may be due to other variables like extent of singing exposure and experience, and that the singing voice may deteriorate over time if not maintained through engagement.

The findings from Goetze, Cooper, and Brown (1990), Green (1994), Rutkowski and Miller (2003a), Hornbach and Taggart (2005), Rutkowski (2015), and Demorest and Pfordresher (2015) may imply that teaching students to use recording technology to provide aural feedback may help them sing more accurately. If modeling, as Goetze, Cooper, and Brown (1990) outlined, is an essential part of vocal development, then studying the effects a digital recording of a quality child vocal model has on a child's vocal development seems logical. Since using recording technology would involve solo singing, findings by Green (1994) that participants seemed much more comfortable singing together in a group than singing alone would support investigating whether or not singing with recording technology can be used as an aid for children to develop their voices. Rutkowski and Miller's (2003a) finding about frequency of instruction support investigating whether or not children can self-listen and make improvements without multiple listenings. The findings from Hornbach and Taggart (2005), and Demorest and Pfordresher (2015) suggest that because neither musical aptitude or age are clear

indicators of pitch accuracy, another explanation of pitch accuracy is needed.

Rutkowski's (2015) findings show that pitch accuracy is directly connected to use of singing voice, which may be useful in understanding how children approach listening back to their own voice.

All of the studies so far have utilized human judges to provide the data for pitch accuracy. As previously noted by Goetze, et al. (1990) these types of studies are not as reliable as those using digital technology to assist in pitch-matching assessment. The following two studies incorporated technology and/or judges for assessment.

Both, technology and human judges were used in Hedden and Baker's (2010) study of accompaniment feedback. Hedden and Baker (2010) examined how well second-grade children matched pitch while singing a newly learned song a cappella or accompanied. The authors stated that previous literature showed conflicting findings concerning how children best sing accurately and inquired (a) in which condition, singing a cappella or with accompaniment, is children's singing more accurate? (b) Did learning condition, that of singing a cappella or with accompaniment, demonstrate a difference in children's singing accuracy? and (c) Is there a difference between acoustical and perceptual analyses of the children's singing? (p. 39).

Participants were 26 second-grade students from two elementary schools' music classes in a Midwestern city. Researchers spent time in the classroom before the study to familiarize themselves to the students. The same repertoire, resources, and rehearsal procedures were maintained between both schools. After practicing the tune several times with the teacher over two class periods, students were asked to sing "America" on a

neutral syllable “loo” so that lyrics would not interfere with pitch-matching. Singing was recorded twice, once a cappella and once with recorded piano accompaniment through headphones. Data were analyzed for pitch accuracy to 50 cents, and perceptually via three human judges marking incorrect pitches. The inter-judge reliability ranged from .86 to .88.

Hedden and Baker (2010) also examined note accuracy based on gender and found no significant difference. There was also no significant difference in the singing accuracy between children who learned the song with accompaniment or without. Judges' analyses indicated a statistically significant ($p < .001$) finding that students sang more accurately when accompanied than unaccompanied, but acoustical analysis showed no significant difference. The authors suggest that further research investigate whether or not the use of a neutral syllable affects pitch accuracy. These conclusions support investigations into the differences between using real-time vs. reflective (listening to a recorded voice) feedback since recording technology can provide both scenarios.

Using only technology to analyze data, Pfordresher and Brown (2007) conducted a quasi-experiment on poor-pitch singing, hypothesizing that problems with pitch accuracy stem from production, memory, and/or sensorimotor integration. The researchers asserted that previous research of poor-pitch singing has focused too heavily on perceptual deficits, and that more recent research has suggested that there are several different deficit models: perceptual, motor, imitative, and memory. Pfordresher and Brown (2007) designed two experiments to test their hypothesis that predicted pitch

inaccuracy to be part of the memory deficit model, and to test the assumptions of the motor deficit model.

Participants in the first experiment were 79 volunteers from a psychology class from a major university in Texas who reported having no formal musical training and no vocal pathology or hearing issues. The experiment consisted of production trials which included a guided warmup, production tasks such as echoing, and a discriminating perception task. Data were collected on pitch accuracy for single pitch and a pitch sequence with both “normal feedback” where the participant heard his or her own voice through headphones, and “augmented feedback” where the participant heard a synthesized voice through the headphones. Pitch accuracy was analyzed for both note accuracy and interval accuracy using computer software. Participants whose note accuracy was on average 100 or more cents (a semitone) off in either direction were considered to be poor-pitch singers. Data collected via questionnaire included demographic information, beliefs about personal singing and musicality, and information about their past exposure to music and singing (Pfordresher & Brown, 2007).

In the first experiment, 13% of the participants were identified as poor-pitch singers. Participants identified as poor-pitch singers consistently “transposed” their produced pitches either consistently sharp or flat. To further test the reliability of this finding, Pfordresher and Brown (2007) compared the “transpositions” by key area, and the data confirmed that participants identified as poor-pitch singers “transposed” toward the key area and did not center on a singular pitch area. There was also a statistically significant ($p < .01$) difference between the interval distance accuracy of poor-pitch

singers and on-pitch singers, where the data indicated that poor-pitch singers compress their intervals more than on-pitch singers. Pfordresher and Brown (2007) also examined mean note errors with normal, augmented, and masked feedback. The statistically significant ($p < .01$) mean difference from this examination indicated pitch-matching singers performed less accurately as task complexity increased, but performed significantly better when singing with augmented feedback. Contrastingly, poor-pitch singers matched pitch more accurately as complexity increased, and performed more accurately under normal feedback. Lastly, Pfordresher and Brown (2007) found no significant differences in pitch discrimination between poor-pitch singers and on-pitch singers.

Pfordresher and Brown (2007) suggest that their findings concerning singers who were able to imitate pitch within 100 cents but unable to match within 50 cents, paired with the data showing no significant difference in pitch discrimination ability indicated poor-pitch singing may be an issue of interval compression. Another finding suggests that poor-pitch singers perform less accurately with accompaniment or correctly modeled than they do only listening to themselves, and more accurately as the complexity increases. Pfordresher and Brown's (2007) study generates some new questions about how pitch-poor singers and on-pitch singers internalize their own singing. Their findings support the notions that pitch-poor singers might benefit from listening back to their own recorded voice, whereas on-pitch singers may benefit from listening back to a model.

Looking lastly at feedback and modeling as it pertains to singing accuracy, Rutkowski and Miller (2003b) examined how teacher feedback and modeling affects the

singing accuracy of first graders. The authors stated that anecdotal literature stresses the importance of teacher feedback and modeling in developing greater singing accuracy in students, but the two had not been empirically investigated. Rutkowski and Miller (2003b) examined aptitude scores using the IMMA (Gordon, 1986) and student singing ability using SVDM.

Participants were 38 first graders from a single elementary school music class in Pennsylvania who had been introduced to group and individual singing practices. The children in the treatment group received specific verbal feedback after singing, and teachers would use modeling to show both inaccurate singing and exemplary singing juxtaposed for the student to hear. Children in the non-treatment group were not given feedback such as “good” or “not bad” and the teacher did not give further instruction beyond the modeling. Rutkowski and Miller (2003b) administered both the IMMA and the SVDM prior to the treatment and after the treatment, and the IMMA midway through the treatment. The SVDM was recorded on tape and scored by two judges with inter-rater reliability scores ranging from .85 to .94.

Musical aptitude indicated that the only statistically significant difference ($p < .001$) was from the IMMA midtest (14.61) to the posttest (18.28). Singing ability as measured by the SVDM showed an increase in singing ability on text in the treatment group (45%) greater than that of the control group (22%), and on “bum” in the treatment group (55%) greater than the control group (28%).

Rutkowski and Miller (2003b) suggest that the lack of differences between aptitude tests may be due to outside factors, and suggest that teacher modeling in the

large group setting does not have a significant impact on student singing. The authors do propose that students may benefit more from hearing their own voice which may be as beneficial to that of teacher feedback, and that further research investigating effective strategies regarding modeling and feedback should be investigated (p. 8).

Based on this literature provided by Pforshresher and Brown (2007), Hedden and Baker (2010), and Rutkowski and Miller (2003b), investigating using a recording device as a self-monitoring and modeling strategy for helping children to develop an accurate singing voice seems necessary.

Theoretical and Practical Aspects of Musical Self-Regulation

Self-regulated learning (SRL) has been investigated for the purpose of understanding the student's involvement in his or her own learning metacognitively, behaviorally, and motivationally. Inquiries into self-regulation in music have been focused on motivation and practice habits (e.g. Zimmerman, 2000). Literature in the general education field pertaining to the development of self-regulatory learning (SRL) has provided theoretical models for how SRL functions. Green and Azevedo (2007) reviewed several theoretical models and focused on Winne and Hadwin's (1998) four phase model of SRL. In this four phase process students experience a cyclical process involving task definition, goal setting and planning, studying tactics, and adaptations to metacognition. Green and Azevedo (2007) asserted that this model is comprehensive, integrates various theories of SRL, and provides a scaffolded model for how SRL is developed by explaining how learners move through each phase. The authors suggest

that research investigate how these phases work at various age levels to help clarify how SRL develops over time (p. 365).

In *The New Handbook of Research on Music Teaching and Learning*, McPherson and Zimmerman (2002) outline important aspects of highly self-regulated learners and aspects of SRL that are of interest for further research. The authors pointed to the development of task-oriented strategies as a crucial aspect of successful student practice habits. Mental strategies (e.g., pre-performance reminders) and self-instruction strategies (e.g., defining musical goals rather than technical goals) were also discussed as strategies that self-regulated learners employ. McPherson and Zimmerman (2002) stated that self-regulated learners employ forethought strategies that are highly reflective and, therefore, emphasized the importance of metacognition, self-evaluation, and reflection. Modeling and feedback are listed as effective strategies for developing self-reflection, and the authors suggest that further research investigating how and when SRL habits are developed is important. Applications of such research could help students develop the motivation for life-long music learning and participation.

Researchers have looked at strategies based on theoretical SRL models to test how effective these strategies are in terms of student musical development. Bonneville-Roussy and Bouffard (2015) examined how the use of an integrated model of self-regulation and deliberate practice strategies predict musical achievement in instrumentalists. The authors posited that “formal practice” (goal-directed, focused practice) will have a positive effect on musical achievement. Previous literature suggested that the majority of practice time used by students can be defined as “informal”

and thus the authors saw an opportunity to investigate the SRL elements of formal practice.

Participants were 173, 16 to 30-year-old musicians with a variety of different musical backgrounds and specialties. Data on 12 variables (age, experience, self-perceptions of competence, weekly practice, weekly work-play, goal direction, focused attention, self-regulation strategies, deliberate practice strategies, final grade in music, and musical achievement) were collected. The preliminary analysis revealed a significant ($p < .05$) negative correlation between student age and deliberate practice strategies. Other findings suggested a significant ($p < .001$) correlation between self-regulation and weekly practice ($r = .45$), deliberate practice strategies and goal direction ($r = .47$), focused attention and self-regulation strategies ($r = .41$), and deliberate practice and self-regulation strategies ($r = .47$). Bonneville-Roussy and Bouffard (2015) suggested that these findings help to define “formal practice” in terms of goal direction, focused attention, self-regulation strategies, and deliberate practice strategies. They conclude that “in order for practice time to be focused and efficient, musicians need to have prior knowledge of the areas in which they need improvement” (Bonneville-Roussy & Bouffard, 2015, p. 698).

Such findings support the idea of investigating whether or not recording devices used to determine gaps in one's singing accuracy would help foster self-regulation via formal practice. In a dissertation, John Robert Zimmerman (2005) examined relationships between self-listening and recording. Zimmerman (2005) was interested in self-recording, self-listening, and self-evaluation with high school instrumentalists as it

pertains to motivation and self-esteem in music. In reviewing the literature for the use of recording devices as a way for self-reflective listening, Zimmerman (2005) was “surprised to find a scarcity of studies employing tape recording and self-evaluation” (p. 5) and stated self-reflective listening is an “essential” (p. 5) tool for making musical progress. He concluded that investigating the use of recording devices as a way for students to measure their own progress was important.

Participants were 93 high school woodwind instrumentalists in a single large midwest high school. The group was randomly divided into three groups (Listeners-Treatment, Non-listeners Treatment, and Control) and preliminary data on grade level, gender, and ensemble placement/ability level were gathered. The study was conducted in three phases. In phase one all of the students completed a self-report inventory to reveal factors of motivation and self-esteem. In phase two, students in the listening group recorded 60 to 90 seconds of the first nine lessons and completed instructor-designed self-evaluation requiring them to set goals as well. Students in the non-listening group did not listen or self-evaluate during the first nine lessons. In the tenth lesson both treatment groups listened back to their entire ten lessons of recordings. The control group did not record or self-evaluate. In phase three, after the treatment period, all participants once again completed the same self-report inventory. Results from this study showed a statistically significant ($p \leq .05$) improvement from the pretest and posttest in self-perception scores in the listeners group. No other significant results were found.

Zimmerman (2005) suggested that this indicated that “systematic self-listening and self-evaluation created a positive enhancement of student perception of progress of

musical ability” (p. 98). Furthermore, Zimmerman (2005) described a need for further investigation of how recording tools can be used in the classroom as a self-listening and self-reflective tool.

Opportunities for investigating how self-regulatory behavior develops in children seem plentiful when examining using recording technology with self-listening. In order to better understand if self-listening will be an effective self-evaluation technique, it will be important to explore additional research dealing with musical self-perception, self-recognition, and self-awareness in children.

Self-Perception, Self-Recognition, and Self-Awareness

As there is no known research investigating the use of recording devices for self-listening with young children's singing voices, one issue that could arise in self-regulation, self-listening, and self-evaluation is an inability to recognize one's voice and therefore an inability to make adjustments. Current research in speech and language pathology suggests that children can accurately identify their own voices.

Strömbergsson (2012) studied if young children, both with phonological impairment (PI) and without, could accurately identify their own speaking voices. The research questions included:

Do 4-5-year-olds and 7-8-year-olds perform on the same level when identifying their recorded voice as their own? Do children with PI perform on the same level as peers with typical speech and language when identifying their recorded voice as their own? Is the children's performance dependent on the time interval between recording and play-back (immediate vs. delayed by 1–2 weeks)? Do

children with PI use their speech deviance as a cue to identify their own recordings, that is, do they recognize their own recordings better if the recordings are produced with deviant speech? (p. 36)

In the study 48 children ages four to eight with normal speech development were ask to mimic the words spoken by a recorded adult voice on a computer which then recorded their response. The child's recording was then presented with three other children's voices of the same word. The participant would then select on the computer which voice was his or hers by pointing at the screen. Results revealed that 93% of the participants with normal hearing performed with more than 90% accuracy.

Strömbergsson (2012) suggested that the use of "recordings of the child's own speech might well be used in phonological intervention" (p. 42). Strömbergsson's (2012) findings and conclusions suggest that using recording devices could be useful in helping children identify errors in their own voices.

Self-Evaluation, Rubrics, and Technology

Strategies for self-regulation include goal setting and self-evaluation which could be aided with self-assessment techniques that incorporate rubrics. The following studies examine the importance of these ideas, including using technology, as they pertain to developing self-regulatory skills.

Hewitt (2011) investigated whether self-evaluation instruction had an effect on self-evaluation accuracy and music performance accuracy among middle school band students. Hewitt asserted that self-evaluation is an important aspect of self-regulation and has been shown to have a positive impact on student learning across multiple

disciplines. Previous studies (e.g. Bergee, 1993; Byo, 1990; Kostka, 1997) suggested that middle school music students are inadequate self-evaluators. Therefore, the effect of self-evaluation instruction had on self-evaluation accuracy was questioned.

Hewitt (2011) studied 211 middle school band students (grades 5-8) from a highly diverse metropolitan mid-Atlantic state school. Students from this school were required to participate in band. Student self-evaluations and teacher evaluations were scored using the Woodwind Brass Solo Evaluation Form. Each grade level was split into three self-evaluation categories: self-evaluation instruction, self-evaluation only, and no self-evaluation. A four-step process for teaching self-evaluation was employed which included “(a) defining the criteria for evaluation, (b) learning to apply the criteria to their own performances and those of others, (c) receiving teacher feedback on their self-evaluations, and (d) engaging students in the development of goals and action plans to implement in their learning” (Hewitt, 2011, p. 10). A pretest and posttest were given surrounding an eight-week period during which there was a minimum of 500 minutes of instruction time. Performance pieces were selected from solo works and instrumental method books unfamiliar to the students. During the first week of instruction the students participating in the self-evaluation instruction group helped design the evaluation rubric with the help of their teachers.

The data collected by Hewitt (2011) were analyzed using a general linear model with repeated measures and multiple dependent variables to determine the differences among the three self-evaluation groups and four grade levels from pretest to posttest on 24 variables. The three-way interaction for group x grade x time, the two-way

interactions for grade x time, and the followup univariate effect of self-evaluation of tone were found to be statistically significant ($p < .05$). There were no differences in the means of groups for performance technique/articulation or tempo accuracy, and the scores for other dependent variables remained either constant or showed no statistically significant differences.

Hewitt (2011) concluded that instruction in self-evaluation may have little impact on middle school band students' self-evaluation accuracy. Hewitt (2011) stated that it is difficult to figure out why the students were so inaccurate in self-evaluating and that asking students to self-evaluate too many categories simultaneously could be an issue. Hewitt (2011) suggested focusing on a single component using only rhythm or melody self-assessments as these categories have been most accurate in previous studies. Hewitt (2011) also attributed a lack of growth in self-evaluation accuracy to the loss of practice time, concluding that musical experience and cognitive development play a role in self-evaluative skill. Although Hewitt (2011) concluded that students were poor self-evaluators, the missing component from his study was the self-listening element. Self-assessments in general elementary school classrooms have been cited as more effective when paired with rubrics (e.g. Bingham, Holbrook, & Meyers, 2010) and, perhaps most critically, self-reflection through review of work (i.e. editing one's writing). The inaccuracies of musical self-evaluation may be based more in the child's inability to simultaneously perform and self-evaluate.

Most pertinent to the present study is the use of technology as an aid in the evaluation process. In one such study Paney and Kay (2014) examined the effects of

using computer games on pitch-matching abilities in third-grade students. The authors postulated that because previous research (e.g., Welch et. al., 2009) suggests that students would benefit from a pedagogical approach to singing that narrows the gender gap, increases enthusiasm, gives specific pitch-matching feedback, and allows for individual growth, using computer games that target these areas was worthy of investigation (Paney & Kay, 2014, p. 3). The research questions posed were:

- (a) Will pitch-matching scores improve with immediate, concurrent visual feedback?
 - (b) Will there be a difference in scores for boys and girls?
 - (c) Will students' scores improve with more time between pretests and posttests?
 - (d) Will students' scores improve with more practice sessions?
 - (e) Will there be a difference in scores based on the key of the example?
 - and (f) Will there be a difference in scores based on the visual layout?
- (p. 3)

Participants were 2,021 third-grade students at 30 different elementary schools in four midwest states. The pretest and posttest consisted of participants singing "America" into a microphone while watching their sung response on the computer screen via either a continuous scrolling line or traditional notation. The program used, *SingingCoach*, would then give the participant a score out of 100 possible points based on the percentage of correctly sung pitches and note attacks.

After analyzing the data by a five-way between groups univariate analysis of variance, Paney and Kay (2014) found a statistically significant ($p < .001$) difference between participant's mean pretest (46.83) and posttest scores (56.99), and also found that using both visual formats was more beneficial than using either alone. The

researchers also noted that students who only used the program during the pretest and posttest had a significantly smaller increase in mean score (4.70 points of growth) than did participants who also used the program one to five times in between the pretest and posttest (13.73 points of growth).

Paney and Kay (2014) conclude that the significant improvements made using the computer program as habit is beneficial in helping students develop and strengthen pitch accuracy and singing ability. The researchers highlighted the importance of differentiated learning and how the use of combined visual formats (both scrolling and notation) may have helped deepen the understanding and relationship between the voice and the visual aspects of music. Paney and Kay (2014) also noted that there was an extensive amount of positive feedback about how the participants enjoyed using the program, including students who had struggled to match pitch in the past. Another provocative conclusion from this study is the “lack of significant difference between the scores of third-grade girls and boys” (p. 7) which Paney and Kay (2014) attributed to the game element the software provides, which might be an increased motivator for boys.

Paney and Kay's (2014) study poses many important questions for the current study about the nature of using technology for evaluation. First, is the use of a visual element key to children being successful in making progress in pitch-matching? Second, do the students only benefit from having their scores given to them by the program, or can they be taught to self-evaluate their own singing? Last, does the game element of using technology have a greater motivating effect, or does simply using recording technology act as a sufficient motivator?

Summary

The literature covered in this chapter provides a framework for understanding research that has examined singing accuracy, self-regulation, self-listening and self-perception, and self-evaluation. Important conclusions from the literature include that students benefit most from the use of a child model (Geotze et. al., 1990; Rutkowski & Miller, 2003b), the singing voice is malleable and can be strengthened through practice (Hornbach & Taggart, 2005), self-evaluation is best employed when goal setting and rubrics are employed (Bonneville-Roussy & Bouffard, 2015; Hewitt, 2011), and technology can successfully be used to assist children's singing voice development (Paney & Kay, 2014)

The current study focused on the following unanswered questions: "Will children's melodic-singing accuracy benefit by using recording technology and self-listening to self-assess?" and "Will students be able to build self-regulatory skills through self-listening to both recordings of themselves and a model?" Findings from current research suggest that recording technology may be a benefit through the use of a digitally provided model voice, provided feedback that doesn't rely on the child's ability to process in real time, and development of skills for self-regulation and meaningful self-evaluation. The following study was conducted to determine if these topics are relevant to voice development, self-evaluation, and self-regulation skills in second-grade students using recording technology.

Chapter 3:

Methodology

The first two chapters presented an overview of research related to children's singing ability, self-regulatory and self-evaluative skills, and the use of technology as an aid for both. The purpose of this study was to investigate the effects of self-evaluation, self-recording along with self-listening, and modeling on second-graders' melodic singing accuracy.

Participants

Participants were 49 second-grade public school children (ages 7 to 9; girls, $n = 28$ and boys, $n = 21$) from two suburban schools in one school corporation in southern Indiana (School 1, $n = 39$ and School 2, $n = 10$) who attended a general music class twice a week for 30 minutes with a licensed music specialist. Volunteer participants from these schools were randomly assigned to either the control group or one of the two treatment groups. Participants were assigned an ID number carrying only their school, gender, self-rating response, and treatment group information.

Measures

The measures used in this study included a self-evaluation rubric and melodic singing accuracy analyses of each students' recorded singing using the software PRAAT. The self-evaluation rubric was a researcher-adaptation of the Singing Voice Development Measure (SVDM) (Rutkowski, 1990a). The SVDM was altered so that it employed age-appropriate language. In addition, the five levels of the SVDM (see Appendix B) were re-

worded in the rubric so that participants were able to use it as a generic framework for self-evaluating their singing voice (see Figure 1).

- 1- I used my talking voice
- 2- I tried to use my singing voice, but it sounds like my talking voice
- 3- I used my singing voice but didn't get most of the notes
- 4- I used my singing voice and got most of the notes
- 5- I used my singing voice and got all of the notes

Figure 1. Adapted SVDM: Version 2 Rubric.

Participants' singing was recorded on a MacBook Pro using GarageBand. His/her voice was captured with a condenser microphone in the "recording booth" (see Appendix D for photographs of the recording booth). The participants listened back to their recorded singing via headphones. The recorded voice files were analyzed for pitch accuracy using PRAAT by finding the mean frequency (in hertz) for each note sung. The entire sound envelope of each pitch was analyzed. Steady-state pitch measurement was not used because only measuring the sung pitch's steady state would not provide a holistic picture of the participant's melodic singing accuracy. Rather, only extreme fluctuations caused by outside noise or vocal plosives were removed from the mean frequency analysis. A melodic singing accuracy score was calculated as the adjusted deviation from the frequency of each pitch and was measured as absolute value of cents from pitch center in equal temperament tuning and A = 440 Hz (Pfordresher & Brown, 2007).

The participants were asked to sing the song “Busy Buzzy Bee” during the experiment (see Figure 2). “Busy Buzzy Bee” was chosen because it required the participant to access the head voice to match pitch, is written in a developmentally appropriate range (D4 to B4), and featured a melody that can be sung using the solfege syllables Sol, La, Mi, Fa, Re, and Do. “Busy Buzzy Bee” was also chosen because of the percussive consonants which allowed the PRAAT analysis tool to identify the beginnings and endings of pitches more easily.



Figure 2. “Busy Buzzy Bee”.

Treatment

The experimental conditions consisted of two treatment groups and a control group. Participants were randomly assigned to the control group or one of the treatment groups. The two treatment groups and the control group all had access to a singing rubric (see Appendix B) which was introduced and used in the music classrooms two class periods prior to the treatment. Participants in the control group (a) sang the song, (b) used the rubric to rate themselves without listening back to any reference recordings, and (c) sang the song again. Participants in treatment group A (a) sang the song, (b) used the rubric to rate themselves, (c) were asked by the researcher to listen back to their own

recording, and (d) sang the song again. Participants in treatment group B (a) sang the song, (b) used the rubric to rate themselves, (c) were asked by the researcher to listen back to their own recording, and then (d) were asked by the researcher to listen to a recorded model, and (e) sang the song again.

The recorded model used in treatment B was of a second-grade girl singing the song. The original recording of the voice was altered via pitch corrective software to achieve an accurate yet natural sound. The final pitch-altered version of the recording resulted in 94.32% pitch accuracy.

A set of prompts was used as a guide to deliver instructions to participants during the treatment (see Figure 3). The participant sang alone behind a partition referred to as the “recording booth” with the microphone and rubric, while the researcher remained on the other side of the partition. A lyrics sheet was present on the recording booth wall for the participant to reference.

Control Group Prompt

- Sing and I will record your voice
- Rate yourself with the rubric
- Sing again and I will record your voice

Treatment A

- Sing and I will record your voice
- Rate yourself with the rubric
- Listen to yourself as I play it back
- Sing again and I will record your voice

Treatment B

- Sing and I will record your voice
- Rate yourself with the rubric
- Listen to yourself as I play it back
- Listen to this model singing
- Sing again and I will record your voice

Figure 3. Set of prompts read to participants.

Procedure

The researcher taught all students to use the singing rubric during two usual class meetings prior to the treatment. The researcher introduced the song “Busy Buzzy Bee” (see Figure 1) by singing the song in falsetto for the students. The students also sang the song by themselves during two class meetings prior to the treatment using a scripted song teaching procedure (see Appendix C) and reviewed the song again during

the class period one day prior to the treatment. An immersion/holistic approach (as opposed to an echo/pattern approach) was used to teach the song (Klinger, Campbell, & Goolsby, 1998).

A space was prepared in the music classrooms at each school in which the participant and the researcher could be away from any extraneous noise but still in full view of the music teacher. The participant stepped into the “recording booth” (see photographs in Appendix D) which acted as a visual partition between the researcher and the student but allowed the researcher to monitor what the student was listening to and recording. The researcher read the prompt (see Figure 3) and proceeded to record the participant’s session using the condenser mic and GarageBand. Participants in treatment groups A and B were prompted by the researcher to listen back to their initial recordings of the song. Participants in treatment group B were also prompted when it was time to listen to the model recording. Recordings of the participants’ sessions were labeled with numerical codes to indicate each participant’s school, group, and self-evaluation response, and were then imported to a computer to be analyzed for pitch accuracy.

Analysis

The dependent variable (melodic singing accuracy) was analyzed using PRAAT pitch accuracy software by extracting the mean frequency of each sung note. Deviations in Hertz for each note sung were calculated and transformed into cents using:

$$\text{note deviation} = 1,200 * \text{Log}_2(\text{sung note in Hz} / \text{correct note in Hz})$$

Since none of the participants sang the song in the key of D where it was written, the follow steps were taken to adjust the pitches sung by the participant to the key center of

D. An average (in Hertz) of all 14 “A’s” sung in the piece was calculated for each recording. This “A” average was used as the reference point from which to pitch shift the recording to A = 440. The difference in Hertz between the “A” average and A = 440 was transposed into an average deviation in cents using:

$$\text{participant's deviation} = 1,200 * \text{Log}_2(\text{Average "A"}/A = 440)$$

The average deviation in cents was then added to each of the participant's actual deviations to create a set of adjusted deviations. The average of the absolute values of each of the deviations was then calculated yielding the participant's melodic singing accuracy average deviation score. Melodic singing accuracy measurements were made for both pre- and post-treatment recordings of participants' attempted singing (See Appendix E for example of data).

Using a pitch shifted “A” average for analysis was beneficial for measuring the participants' melodic singing accuracy because it allowed for the participant to sing in any key (including keys between equal temperament pitches). Using an average instead of a single note as the fundamental note of the pitch shift maintained the integrity of the center of tonality by not favoring the first note (as is with pitch shifting via a DAW program like Audacity). Using only the “A’s” in the song removed biases towards participants who sang both equally sharp and flat over participants who sang only sharp or flat by averaging the most common single note rather than the average of the entire melodic contour.

Chapter 4:

Results

Overview

The purpose of this study was to investigate the effects of self-evaluation, self-recording along with self-listening, and modeling on second-graders' melodic singing accuracy. Treatments were designed to evaluate whether variation in melodic singing accuracy when using a rubric would vary as a function in the presence or absence of self-listening and presence or absence of a recorded model. The data were collected from 49 participants, but one participant's data was unusable because the participant did not sing recognizable pitches ($N = 48$). The three groups consisted of a control group ($n = 16$), treatment A ($n = 16$), and treatment B ($n = 16$). The data were collected over a three-week period at each school (School 1, December, 2016; School 2, February, 2017).

Analysis

An exploratory analysis of the pre-test, post-test, and growth measures of pitch accuracy revealed that Treatment B contained three outliers (see Table 1). Two of the outlying participants were identified as having trouble maintaining a light voice and were therefore losing pitch accuracy, while the other outlying participant had trouble producing sound in her first recording but used more support in the second. The data for these participants were removed from further analyses. Descriptive analyses without the outliers are presented in Table 2.

Table 1

Descriptive Analysis

| Treatment | <i>M</i> | <i>SD</i> | <i>Skew</i> | <i>Kurtosis</i> |
|----------------------|----------|-----------|-------------|-----------------|
| Pre-Test | | | | |
| Control | 129.26 | 68.51 | .12 | -1.23 |
| Treatment A | 89.97 | 61.31 | .96 | -.23 |
| Treatment B | 92.62 | 60.61 | 1.19 | .92 |
| Post-Test | | | | |
| Control | 139.75 | 75.97 | -.31 | -1.5 |
| Treatment A | 91.59 | 64.59 | .89 | -.61 |
| Treatment B | 84.96 | 62.03 | 1.44 | 1.54 |
| Growth (Pre to Post) | | | | |
| Control | -10.50 | 24.56 | -1.10 | 1.56 |
| Treatment A | -1.62 | 13.08 | -.53 | .50 |
| Treatment B | 7.67 | 56.07 | .19 | 1.48 |

Table 2

Descriptive Analysis without Outliers

| Treatment | <i>M</i> | <i>SD</i> | <i>Skew</i> | <i>Kurtosis</i> |
|----------------------|----------|-----------|-------------|-----------------|
| Pre-Test | | | | |
| Control | 129.26 | 68.51 | .12 | -1.23 |
| Treatment A | 89.97 | 61.31 | .96 | -.23 |
| Treatment B | 79.85 | 43.38 | 1.20 | 1.30 |
| Post-Test | | | | |
| Control | 139.75 | 75.97 | -.31 | -1.5 |
| Treatment A | 91.59 | 64.59 | .89 | -.61 |
| Treatment B | 66.63 | 45.17 | 2.00 | 4.74 |
| Growth (Pre to Post) | | | | |
| Control | -10.50 | 24.56 | -1.10 | 1.56 |
| Treatment A | -1.62 | 13.08 | -.53 | .50 |
| Treatment B | 13.21 | 28.19 | .68 | -.05 |

A 3 x 2 mixed design ANOVA was run to compare average pitch deviation between treatment groups and from pre- and post-test. The test of between subject effects was significant ($p < .05$), indicating that the participants in the control group exhibited significantly more pitch deviation than the participants in the treatments groups overall. Tests of within-subjects effects showed that time alone was not significant indicator ($p = .913$) but that the interaction of time and treatment was significant ($p < .05$) and accounted for 16% of the variance in pitch deviation scores ($\eta^2 = .161$). The interaction revealed that only participants in treatment group B made significant improvement in

pitch deviation (See Figure 4). The data suggest that better pitch-matching accuracy was yielded when a recorded vocal model was present.

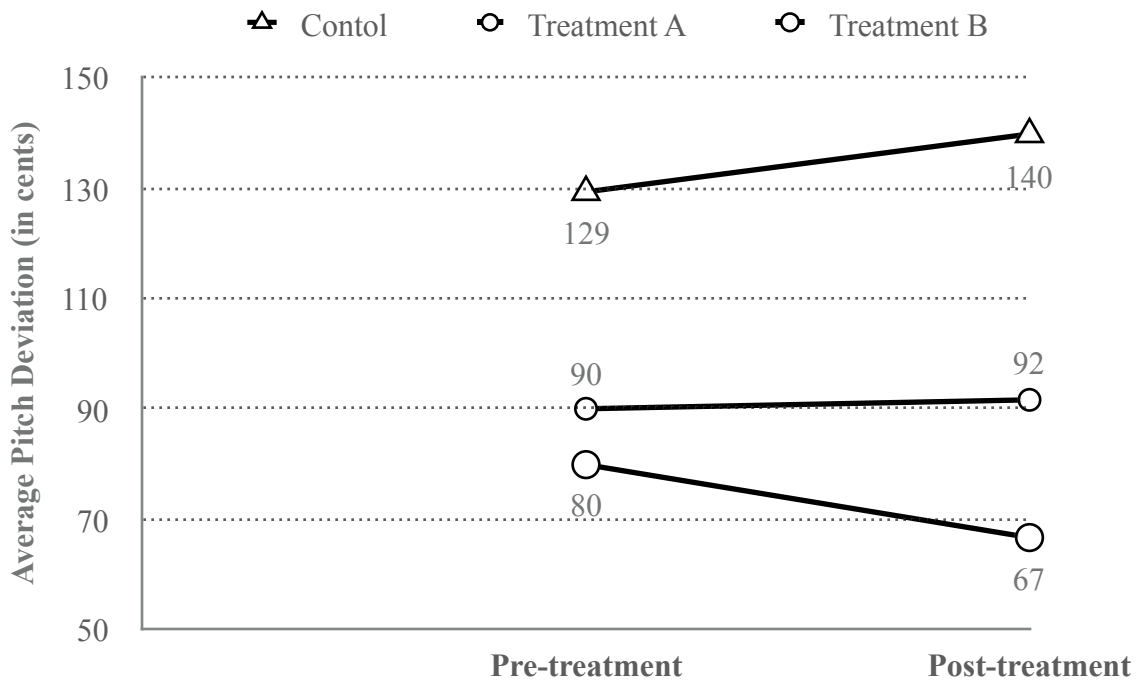


Figure 4. Results of a 3 x 2 mixed design ANOVA of treatments and recordings.

Variation in pitch deviation scores according to gender and treatment was also analyzed using a 3 x 2 ANOVA (e.g., treatment by gender). The interaction was non-significant ($p > .05$) showing that gender did not interact with the effect of the treatment.

Unexpectedly, several participants sang the wrong melody either one or both times ($n = 18$). Since the data were analyzed using the specific pitches from the song as a reference point, singing a different melody would result in severe deviations no matter how well it was executed. Eight of the participants from the control group, seven from treatment group A, and three from treatment group B sang the melody incorrectly at pre-test. This is a partial explanation for the high pitch deviation scores of the control group at pre-test. Curiously, no participants from the control group who sang an incorrect

melody at pre-test were able to correct their singing at post-test. However, some participants from treatment groups A and B who sang an incorrect melody at pre-test were able to correct their singing at post-test (see Table 3).

Table 3

Crosstabulation of Melody Accuracy per Treatment Group

| Treatment | Correct Both Times | Incorrect First and Corrected Second | Incorrect Both Times |
|-------------|--------------------|--------------------------------------|----------------------|
| Control | 8 | 0 | 8 |
| Treatment A | 9 | 4 | 3 |
| Treatment B | 13 | 2 | 1 |
| Total | 30 | 6 | 12 |

Another unexpected finding was that many of the participants ($n = 10$) switched from using a “heavy voice” at pre-test to a “light voice” at post-test (Goetze, Broeker, & Boshkoff, 2011). Although there is no mathematical delineation between the heavy and light voice due to the frequency crossover [“heavy voice” (A3 to A4) and “light voice” (D4 to F5) (Goetze, Broeker, & Boshkoff, 2011, p. 75)], the timbral difference between the two voices is easily perceptible. The data revealed that none of the control group participants switched from heavy to light voice, but some participants from both treatment groups did (see Table 4).

Table 4

Crosstabulation of Heavy to Light Voice Switch per Treatment Group

| Treatment | No Change | Switch to Light Voice | Total |
|-------------|-----------|-----------------------|-------|
| Control | 16 | 0 | 16 |
| Treatment A | 14 | 2 | 16 |
| Treatment B | 8 | 8 | 16 |
| Total | 38 | 10 | 48 |

As a final point of analysis, a Spearman correlation coefficient was calculated to assess the relationship between students' self-ratings and their pitch deviation. No statistical significant correlations ($p > .05$) were found. There was a significant ($p < .001$) negative correlation between the researcher's rubric ratings and each participant's pre-treatment recording (i.e. the higher the rubric rating, the lower the pitch deviation). The correlation seemed low ($r = -.542$) due to the ratings from participants singing the incorrect melody mostly receiving a 3 on the rubric (i.e. used singing voice but didn't get most of the notes) and also receiving a high deviation score (see Table 5).

Table 5

Participant and Researcher Rubric Ratings, and Pre-treatment Deviation Correlations

| | Pre-treatment | Researcher Rating |
|--------------------|---------------|-------------------|
| Participant Rating | -0.059 | 0.264 |
| Researcher Rating | -.542** | |

** correlation significant to the 0.01 level (2 tailed)

Chapter 5:

Discussion

Summary

This study investigated the effects of self-evaluation, self-recording along with self-listening, and modeling on second-graders' melodic singing accuracy. The purpose was to examine the effects of the use of only a rubric for self-evaluation on vocal pitch accuracy (control group), the use of a rubric and self-listening on vocal pitch accuracy (Treatment group A), and the use of a rubric, self-listening, and listening to a recorded model on vocal pitch accuracy (Treatment group B). Forty-nine participants from two elementary schools in southern Indiana were randomly assigned to the three groups and asked to sing a song they had previously learned in class once before the treatment and then again after the treatment. Each participant's attempt at singing both pre- and post-treatment were recorded and the data were analyzed for average pitch deviation (in cents) from the correct notes. The data showed that on average only participants in Treatment B improved their pitch accuracy from pre- to post-test. Though non-significant, the control group on average had slightly worse pitch accuracy at post-test and participants in treatment group A showed neither improvement nor decline at post-test.

The findings on melodic singing accuracy support those from previous studies which suggested that a child model improves pitch accuracy (Goetze, Cooper, & Brown, 1990; Pfordresher & Brown, 2007) and that children can make improvements without practice or multiple repetitions (Rutkowski & Miller, 2003a). The data also suggest that second-grade children may be able to use a model to develop self-instruction strategies

(McPherson & Zimmerman, 2002) to address their own inaccuracies in singing accuracy. The participants in Treatment B may have been able to recognize the difference between the model and their own singing without prompting from the teacher.

The data also lend support to the idea that self-listening and/or listening to a model supports the development of self-regulatory strategies since some participants in Treatment A and Treatment B switched to singing with a “light voice” when they had been using a “heavy voice” ($n = 10$) (Goetze, Broeker, & Boshkoff, 2011), while others in those two groups who sang the incorrect melody the first time sang it correctly the second time ($n = 6$). Since the participants were not given time to practice, these findings suggest that the children may have been able to self-evaluate by making comparisons with the model and then adjustments to their voices even without explicit instruction to do so. Although this finding goes against the suggestions of other previous research that stressed the importance of self-regulated practice as an integral aspect of successful musicianship (Ho, 2014; Kenny, 1998; Kostka, 1997; Wells, 1998), it is congruent with Bonneville-Roussy and Bouffard's (2015) conclusions that “musicians need to have prior knowledge of the areas in which they need improvement” (p. 698). The aid of the recording device for self-listening further supports previous findings that self-listening can be used to identify areas in need of improvement (Pfordresher & Brown, 2007; Silveira & Gavin, 2016; Zimmerman, 2005).

The results of this study are also consistent with findings from previous research suggesting that there are no inherent differences in pitch accuracy between second-grade boys and girls (Hedden & Baker, 2010; Welch, et. al., 1997). Although some previous

studies suggest that boys may sing more poorly due to peer pressure (Hornbach & Taggart, 2005), the privacy of the recording booth in the current study may have mitigated that effect.

Findings from the self-rating scores are congruent with previous studies (e.g. Ross, 2006) suggesting that the use of rubrics and self-evaluation strategies are ineffective by themselves. For example, there was not a statistically significant correlation between the participants' self-ratings and their singing quality, nor their self-ratings and the researcher ratings. In some cases, the participant's self-evaluation score may more closely reflect of the child's self-efficacy as a singer than an actual evaluation of her/his voice.

Limitations

The findings from this study are limited by the small number of participants, the fact that the participants were volunteers, and the mostly homogenous population of participants. Each treatment group contained only 16 participants, making it difficult to generalize the findings to a greater population of second-grade children. The children and their guardians who agreed to participate in the study also do not represent the entire population of available second graders from the two schools used in the study, therefore it is possible that these findings are not be generalizable to a broader population.

Generalizability is also limited due to the participants being from the same general geographic location as well as being from similar cultural contexts.

This study was also limited in method by how the participants' singing ability was measured with the particular song that was chosen, the way in which the song was taught,

and the length of the experiment. Since the participants' voices were analyzed only for average pitch accuracy adjusted to the key of D, this leaves out other measurable variables such as note shape or tone. Since only "Busy Buzzy Bee" was used in the study the results may have been influenced by the familiarity with a song based on sol, la, and mi solfege. This might also explain why some of the participants sang a similar sol, la, and mi based melody instead of the correct melody. The structuring of the lessons for teaching and practicing the song also limit the generalizability of the study since the song was taught with immersion (as opposed to echo) and the researcher sang the song in falsetto as a model. The study is further limited since the time between the pre- and post-treatment recordings was only enough time for the researcher to give instructions and for the participant to think about them.

Implications for Further Research

The findings from this study are encouraging and suggest that further research is needed in studying how self-listening, listening to a model, and opportunities for self-regulation can affect children's singing abilities. Examining how self-listening affects self-evaluation when a model is present seems to be an important next step based on the current and previous studies (Silveira & Gavin, 2016). Since the participants were asked to self-evaluate with the rubric only after singing in real-time or listening to themselves, it would be interesting to give them an opportunity again after listening to a model to rate themselves to see if their rating changed. In this study, the model was heard after self-listening occurred, but it is possible that the timing of the model could have an effect on

singing accuracy. As such, exploring presenting a model at different time points during the learning process would be valuable as well.

The recording booth from this study raised interesting questions similar to Green's (1994) about unison versus solo singing. Green (1994) suggested that children sing better in group settings than alone because they are more comfortable blending into the sounds of other children as opposed to being exposed. However, the findings from this study in regards to the effectiveness of Treatment B also suggest that Green's (1994) conclusions may be circumstantial, as all of the participants in this study sang alone in the recording booth. Since no data were collected on participant's singing ability while singing in a group it is not possible to conclude whether or not the participants sang more on pitch in the group or alone in the recording booth. Further research is needed to reveal more about how children feel about the environment (e.g. in class with their peers or alone in a recording booth) in which they are singing alone or with the group and how that affects their ability to sing. Does the recording booth environment change how children sing alone versus how they sing solo in front of their peers and teachers?

More research is needed to explain why some participants ($n = 4$) in Treatment group A were able to correct an incorrectly sung melody after listening only back to themselves and not a model. Intuitively, it is understandable that participants in Treatment group B would correct an incorrectly sung melody after being able to hear the model sing, but it is harder to imagine a reason for the participants in Treatment group A to correct the melody based solely on self-listening. This phenomenon is intriguing, and

further research should be conducted to evaluate whether or not children can not only detect errors but also self-regulate melody correction by only self-listening.

Researchers interested in measuring melodic singing accuracy could benefit from this study. The “A” average analysis method used in this study is a clearly defined empirical measurement that is more objective and reliable than using subjective rater-judgments for melodic accuracy.

Suggestions for Teachers

Although applying the findings of this study directly to classroom use seems premature, there are some aspects that lead to suggestions for the effective use of self-listening and modeling for classroom teachers. First, child models, both live and recorded, seem to be an effective tool based on the findings from this study and as well as others (e.g. Geotze et. al., 1990; Rutkowski & Miller, 2003b). Child models should be used frequently to assist with children who are continuing to develop their singing voices.

Second, self-listening can be an effective tool for helping children develop a sense of their own voice. Using classroom technologies (e.g., iPods, laptops, etc.) for self-listening may allow children to develop clearer self-perception and allow them to self-reflect more easily. This may help set a foundation for self-regulatory practice by providing assistance with the “self-reflective” and “forethought” phases of self-regulation (McPherson & Renwick, 2011) without the intervention of the teacher.

Third, this study supports the idea that self-evaluation without guidance is ineffective. Congruent with previous research (Bonneville-Roussy & Bouffard, 2015; Hewitt, 2011; Hornbach & Taggart, 2005), children do not seem to have the capability to

accurately self-evaluate in real time. Instead of asking students to self-evaluate with help, pairing self-evaluation with self-listening, listening to a model, peer evaluation, and teacher evaluation may be more effective for developing performance skills while also helping students to develop self-evaluative skills.

Last, giving students space within the classroom to experiment with their voices by themselves may be beneficial. Although no empirical data was taken on how each of the participants felt about singing in the recording booth, all of them were able to sing in the booth and many appeared very excited to do so. A few students from School 1 who had returned assent and consent forms expressed disinterest in participating, but a few students from both School 1 and 2 who had not returned assent and consent forms also expressed interest in participating. Whether children would prefer to sing alone, out loud for/with a group, or alone in a recording booth, it is important for teachers to provide space for successful singing. Moreover, pairing the recording booth with recording technology could also provide teachers with assessment opportunities and perhaps help to document evidence of a child's progress in singing.

Conclusions

The goal of this study was to investigate if and how elementary age children could use self-regulatory strategies to improve their melodic singing accuracy given a number of conditions present in general elementary music classrooms. The findings from this study were congruent with many previous studies which suggest that a child vocal model is effective for helping students achieve both a quality singing voice and pitch accuracy and that children struggle to self-evaluate accurately without help. The findings from this

study also suggest that second-grade children are able to self-regulate without practice via self-listening, which is made stronger by pairing self-listening with a vocal model. These findings seem promising for future research concerning how self-regulation develops in children and how teachers might help students acquire such abilities effectively and efficiently.

References

- Asmus, E. (1999). Music assessment concepts. *Music Educators Journal*, 86(2), 19-24.
stable URL: <http://www.jstor.org/stable/3399585>
- Bailey, L. (2006). A study of motivation and self-regulation among high school students. (Doctoral dissertation, Capella University). ProQuest. UMI Number: 3216029
- Bandura, A., Caprara, G., Barbaranelli, C., Gerbino, M., & Pastorelli, C. (2003). Role of affective self-regulatory efficacy in diverse spheres of psychosocial functioning. *Child Development*, 74(3), 769-782. doi: 10.1111/1467-8624.00567
- Bentley, A. (1969). Measurements and development of musical abilities: Some research interests and findings. *Journal of Research in Music Education*, 17, 41-46. stable URL: <http://www.jstor.org/stable/3344182>
- Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education*, 41, 19-27. stable URL: <http://www.jstor.org/stable/3345476>
- Bingham, G., Holbrook, T., & Meyers, L. (2010). Using self-assessments in elementary classrooms. *The Phi Delta Kappan*, 91(5), pp. 59-61. stable URL: <http://www.jstor.org/stable/27755646>
- Bonneville-Roussy, A. & Bouffard, T. (2015). When quantity is not enough: Disentangling the roles of practice time, self-regulation and deliberate practice in musical achievement. *Psychology of Music*, 43(5), 686-704. doi: 10.1177/0305735614534910

- Byo. J. L. (1990). Recognition of intensity contrasts in gestures of beginning conductors. *Journal of Research in Music Education*, 38, 157-163. stable URL: <http://www.jstor.org/stable/3345179>
- Colwell, R., & Richardson, C. P. (2002). *The new handbook of research on music teaching and learning: A project of the Music Educators National Conference*. Oxford: Oxford University Press.
- Cleary, T., Callan, G., Malatesta, J., & Adams, T. (2015). Examining the level of convergence among self-regulated learning microanalytic processes, achievement, and a self-report questionnaire. *Journal of Psychoeducational Assessment* 33(5), 439-450. doi: 10.1177/0734282915594739
- Cuddy, L., Balkwill, L., Peretz, I. & Holden R. (2005). Musical deficiencies are rare. *Annals of New York Academy of Sciences*. 1060, 311–324. doi: 10.1196/annals.1360.026
- Dalla Bella, S. (2015). Defining Poor-Pitch Singing: A Problem of Measurement and Sensitivity. *Music Perception: An Interdisciplinary Journal*, 32(3), 272-282. stable URL: <http://www.jstor.org/stable/10.1525/mp.2015.32.3.272>
- Davis, L. P. (1981). The effects of structured singing activities and evaluation practice on elementary band students' instrumental achievement (Doctoral dissertation, Ohio State University). University Microfilms International.
- Demorest, S. M., Nichols, B., & Pfordresher, P.Q. (2014). The effect of focused instruction on kindergartener's singing accuracy. Manuscript in preparation.
- Demorest, S. M., Pfordresher, P. Q., & Kelley, J. F. (2014, April). School

music participation: Exploring the role of students' self-concept and singing ability. Paper presented at the National Association for Music Education National Conference. St. Louis, MO.

Demorest, S., & Pfordresher, P. (2015). Singing accuracy development from k-adult: A comparative study. *Music Perception: An Interdisciplinary Journal*, 32(3), 293-302. stable URL: <http://www.jstor.org/stable/10.1525/mp.2015.32.3.293>

Eppink, J. A. (2002). Student-created rubrics. *Teaching Music*, 9(4), 28. ISSN 1069-7446

Fernyhough, C. & Russell, J. (1997). Distinguishing one's own voice from those of others: A function for private speech? *International Journal of Behavioral Development*. 20(4), 651-665. doi: 10.1080/016502597385108

Goetze, M., Broeker, A., & Boshkoff, R. (2011). *Educating young singers: a choral resource for teacher-conductors*.

Goetze, M., Cooper, N., & Brown, C. (1990). Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education*, 104, 16-37. stable URL: <http://www.jstor.org/stable/40318355>

Goetze, M., & Horii, Y. (1989). A comparison of the pitch accuracy of group and individual singing in young children. *Bulletin of the Council for Research in Music Education*, 99, 57-73. stable URL: <http://www.jstor.org/stable/40318325>

Green, G. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *Journal of Research in Music Education*, 38(3), 225-231. stable URL: <http://www.jstor.org/stable/3345186>

- Green, G. (1994). Unison versus individual singing and elementary students' vocal pitch accuracy. *Journal of Research in Music Education*, 42(2), 105-114. stable URL: <http://www.jstor.org/stable/3345495>
- Greene, J. & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning. *New Perspectives and Directions Review of Educational Research*, 77(3), 334-372. stable URL: <http://www.jstor.org/stable/4624902>
- Gordon, E. E. (1986). Intermediate measures of music audiation. Chicago: G.I.A. Publications, Inc.
- Gould, O. (1969). Developing specialized programs for singing in the elementary school. *Bulletin of the Council for Research in Music Education*, 17, 9-22. stable URL: <http://www.jstor.org/stable/40317056>
- Hedden, D. & Baker, V. (2010). Perceptual and acoustical analyses of second-graders' pitch-matching ability in singing a cappella or with piano accompaniment. *Bulletin of the Council for Research in Music Education*, 184, 35- 48. stable URL: <http://www.jstor.org/stable/27861481>
- Hewitt, M. (2000). The effects of modeling, self-evaluation, and self-listening on junior high instrumentalists' music performance and practice attitudes. *Journal of Research in Music Education*, 49(4), 307-322. stable URL: <http://www.jstor.org/stable/3345614>
- Hewitt, M. (2011). The impact of self-evaluation instruction on students self-evaluation, music performance, and self-evaluation accuracy. *Journal of Research in Music Education*, 59(1), 6-20. stable URL: <http://www.jstor.org/stable/23019434>

- Ho, P. (2014). *The effects of peer-evaluation on self evaluation skills in the music classroom* (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No.3621775)
- Hornbach, C., & Taggart, C. (2005). The relationship between developmental tonal aptitude and singing achievement among kindergarten, first-, second-, and third-grade students. *Journal of Research in Music Education*, 53(4), 322-331. stable URL: <http://www.jstor.org/stable/3648430>
- Kenny, W. (1998). Thinking critically in the practice room. *Music Educators Journal*, 85(1), 21-23. stable URL: <http://www.jstor.org/stable/3399154>
- Klinger, R., Campbell, P. S., & Goolsby, T. (1998). Approaches to children's song acquisition: Immersion and phrase-by-phrase. *Journal of Research in Music Education*, 46(1), 24-34. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/3345757>
- Kostka, M. J. (1997). Effects of self-assessment and successive approximations on "knowing" and "valuing" selected keyboard skills. *Journal of Research in Music Education*, 45, 273-281. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/3345586>
- Levinowitz, L. M., Barnes, P., Guerrini, S., Clement, M., D'April, P., & Morey, M. J. (1998). Measuring singing voice development in the general music classroom. *Journal of Research in Music Education*, 46, 35-47. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/3345758>

- Mang, E. (2006). The effects of age, gender and language on children's singing competency. *British Journal of Music Education*, 23(02), 161-174. doi:10.1017/S0265051706006905
- McPherson, G. & Renwick, J. (2011). Self regulation and mastery of musical skills. In Zimmerman, B. & Schunk, D. *Handbook of self-regulation of learning and performance*. New York: Routledge.
- McPherson, G. E., & Zimmerman, B. J. (2002). Self regulation of musical learning. In Colwell, R., & Richardson, C. P., *The new handbook of research on music teaching and learning: a project of the Music Educators National Conference* (327-347). Oxford, UK: Oxford University Press.
- Paney, A. S., & Kay, A. C. (2015). Developing singing in third-grade music classrooms: The effect of a concurrent-feedback computer game on pitch- matching skills. *Update: Applications of Research in Music Education*, 34(1), 42-49. doi: 10.1177/8755123314548047
- Pfordresher, P., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception: An Interdisciplinary Journal*, 25(2), 95-115. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/10.1525/mp.2007.25.2.95>
- Phillips, K. H., & Aitchinson, R. E. (1997). The relationship of singing accuracy to pitch discrimination and tonal aptitude among third-grade students. *Contributions to Music Education*, 24(1), 7-22. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/24126943>

- Rosa, Lassonde, Pinard, Keenan, Belin (2008) Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition* 68, 204–214. doi: 10.1016/j.bandc.2008.04.007
- Ross, J. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, 11(10), ISSN 1531-7714
- Rutkowski, J. (1984). Development of a rating scale to assess individual children's use of the vocal instrument. Unpublished manuscript, State University of New York at Buffalo, Buffalo, NY.
- Rutkowski, J. (1986). The effect of restricted song range on kindergarten children's use of singing voice and developmental music aptitude. Dissertation Abstracts International, 4706A. (UniversityMicrofilms No. DA8619357)
- Rutkowski, J. (1989). The problem singer: What does research suggest? *General Music Today*, 2(3), 24-29.
- Rutkowski, J. (1990a). The measurement and evaluation of children's singing voice development. *The Quarterly: Center for Research in Music Learning and Teaching*, 1(1-2), 81-95. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/3345447>
- Rutkowski, J. (1990b). The comparative effectiveness of individual and group singing activities on kindergarten children's use of singing voice and developmental music aptitude. Paper presented at National Biennial In-Service Conference of Music Educators National Conference, Washington, DC.
- Rutkowski, J. (1996). The effectiveness of individual/small group singing activities on

kindergartners' use of singing voice and developmental music aptitude. *Journal of Research in Music Education*, 44(A), 353-368. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/3345447>

Rutkowski, J. (1998). The nature of children's singing voices: Characteristics and assessment. In B.A. Roberts (Ed.), *Sharing the Voices: The Phenomenon of Singing* (pp. 201-209). St. John's, Newfoundland, Canada: Memorial University.

Rutkowski, J. (2015). The Relationship between children's use of singing voice and singing accuracy. *Music Perception: An Interdisciplinary Journal*, 32(3), 283-292. doi: 10.1525/mp.2015.32.3.283

Rutkowski, J. & Miller, M. (2003a). The Effectiveness of frequency of instruction and individual/small-group singing activities on first graders' use of singing voice and developmental music aptitude. *Contributions to Music Education*, 30(1), 23-38. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/24127025>

Rutkowski, J. & Miller, M. (2003b). The effect of teacher feedback and modeling on first graders' use of singing voice and developmental music aptitude. *Bulletin of the Council for Research in Music Education*, 156, 1-10. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/40319169>

Silveira, J. M., & Gavin, R. B. (2016). The effect of audio recording and playback on self-assessment among middle school instrumental music students. *Psychology of Music*, 44, 880-892. doi: 10.1177/0305735615596375

- Strömbergsson, S. (2012). Children's recognition of their own recorded voice: Influence of age and phonological impairment. *Clinical Linguistics & Phonetics*, 27(1), 33-45, DOI: 10.3109/02699206.2012.735744.
- Varela, W., Abrami, P., & Uptis, R. (2014). Self-regulation and music learning: A systematic review. *Psychology of Music*, 1-20. doi:10.1177/0305735614554639
- Welch, G., Sergeant D., & White, P. (1997). Age, sex, and vocal task as factors in singing "in tune" during the first years of schooling. *Bulletin of the Council for Research in Music Education*, 133, 153-160. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/40318855>
- Welch, G. F., Himonides, E., Papageorgi, I., Saunders, J., Rinta, T., Stewart, C., . . . Hill, J. (2009). The national singing programme for primary schools in England: An initial baseline study. *Music Education Research*, 11, 1–22. doi: 10.1080/14613800802699523
- Wells, R. (1998). The student's role in the assessment process. *Teaching Music*, 6(2), 32- 33. ISSN-1069-7446
- Wesolowski, B. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, 98(3), 36-42. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/41433277>
- Whitcomb, R. (1999). Writing rubrics for the music classroom. *Music Educators Journal*, 85(6) 26-32. doi: 10.2307/3399518

- Wise, K. J., & Slobda, J. A. (2008). Establishing an empirical profile of self-defined “tone deafness”: Perception, singing performance and self-assessment. *Musicae Scientiae*, 12(1), 3–26. doi: 10.1177/102986490801200102
- Yarbrough, C., Green, G., Benson, W., & Bowers, J. (1991). Inaccurate Singers: An Exploratory Study of Variables Affecting Pitch-Matching. *Bulletin of the Council for Research in Music Education*, 107, 23-34. stable URL: <http://www.jstor.org.proxyiub.uits.iu.edu/stable/40318418>
- Zimmerman, J. (2005). *Effects of periodic self-recording, self-listening, and self-evaluation on the motivation and music self-concept of high school instrumentalists* (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3192056)

Appendices

Appendix A

INDIANA UNIVERSITY INFORMED CONSENT STATEMENT FOR

THE EFFECT OF RECORDING TECHNOLOGY WITH SELF-LISTENING ON 2ND GRADERS' PITCH-MATCHING ACCURACY

Your child is invited to participate in a research study of pitch-matching accuracy abilities in second-grade children. Your child was selected as a possible subject because s/he attends public school in MCCSC and has general music class regularly throughout the school year. Please read this form and ask any questions you may have before agreeing to allow your child to be in the study.

The study is being conducted by Dr. Peter Miksza and Dylan Fixmer at the Indiana University Jacobs School of Music.

STUDY PURPOSE

The purpose of this study is to examine how the use of recording technology (i.e. an iPod touch) and self-listening (i.e. where a child is given a chance to listen back to themselves) influence a child's ability to match pitch while singing.

NUMBER OF PEOPLE TAKING PART IN THE STUDY

If you agree to participate, your child will be one of 60 subjects from three school in MCCSC who will be participating in this research.

PROCEDURES FOR THE STUDY

If you agree to be in the study, your child will do the following things:

1. Be taught a simple song by their normal music teacher for the purpose of the study.
2. Sing a song alone into a recording device from behind a partition, allowing the child space to record while maintaining clear line of sight by the music teacher.
3. Rate his/herself on a singing rubric used by the teacher in the classroom/ developed by the researchers.
4. Your child may also be asked to listen back to him/herself and record again, and some children will be asked to listen to a model recording of another unknown child and record again.

RISKS OF TAKING PART IN THE STUDY

While on the study, the risks to your child may include embarrassment due to singing alone or listening to one's own voice.

BENEFITS OF TAKING PART IN THE STUDY

There are no direct benefits to your child's participation, however it is reasonable to expect that this will be an enjoyable experience of recording his/her own voice and the opportunity to self-listen which may or may not help his/her pitch-matching ability and singing voice.

ALTERNATIVES TO TAKING PART IN THE STUDY

Instead of being in the study, you may have your child opt out without any negative consequences.

CONFIDENTIALITY

Efforts will be made to keep your child's personal information confidential. We cannot guarantee absolute confidentiality. Child's personal information may be disclosed if required by law. Your child's identity will be held in confidence in reports in which the study may be published. The audio recordings made of your child's voice will be heard by the researchers and analyzed using pitch analysis software. Your child's name will not be attached to any recording made.

Organizations that may inspect and/or copy your child's research records for quality assurance and data analysis include groups such as the study investigator and his/her research associates, the Indiana University Institutional Review Board or its designees, and (as allowed by law) state or federal agencies, specifically the Office for Human Research Protections (OHRP) who may need to access your child's research records.

PAYMENT

Your child will not receive payment for taking part in this study.

CONTACTS FOR QUESTIONS OR PROBLEMS

For questions about the study, contact the researchers, Dylan Fixmer, at (828) 335-5444 or Dr. Peter Miksza at (812) 855-7253.

For questions about your child's rights as a research participant, to discuss problems, complaints, or concerns about a research study, or to obtain information or offer input, contact the IU Human Subjects Office at or 812-856-4242 or 800-696-2949.

VOLUNTARY NATURE OF THIS STUDY

Your child taking part in this study is voluntary. You may choose not to allow your child take part or may leave the study at any time. Leaving the study will not result in any penalty or loss of benefits to which your child is entitled. Your decision whether or not to allow your child to participate in this study will not affect your current or future relations with Indiana University.

SUBJECT'S CONSENT

In consideration of all of the above, I give my consent for my child
_____ to participate in this research study.

I will be given a copy of this informed consent document to keep for my records. I agree to allow my child to take part in this study.

Parent's Printed Name: _____

Parent's Signature: _____ **Date:** _____
(must be dated by the subject)

Printed Name of Person Obtaining Consent:

Signature of Person Obtaining Consent: _____ **Date:** _____

Form date: September 19, 2016

Signature

Date

Indiana University Assent to Participate in Research**THE EFFECT OF RECORDING TECHNOLOGY WITH SELF-LISTENING ON 2ND GRADERS' PITCH-MATCHING ACCURACY**

My name is Dylan Fixmer and I am doing a research study. A research study is a special way to learn about something. I am doing this research study because I am trying to find out more about if children can use iPods to record and listen to themselves to get better at singing. I would like to ask you to be in this research study.

Why am I being asked to be in this research study?

You are being asked to be in this research study because you are a second grader who attends a music class in MCCSC.

What will happen during this research study?

This study will take place during your normal music class. If you participate, you will be asked to sing the song that you learned in class, "Busy Buzzy Busy Bee," by yourself in a recording booth. You will be asked to rate yourself on how well you did, and some of you will also be asked to listen back to yourself. You may also be asked to listen to a recording of someone else sing the song. You will lastly be recorded singing the song one more time, and then you will go back and join music class.

If you want to be in this study, here are the things that we will ask you to do

- Ask your parents if they have signed the consent form for this study.
- Talk to your parents about being in the study.
- Sign the assent form below.
- Return the assent form to your school music teacher.

Are there any bad things that might happen during the research study?

Sometimes bad things happen to people who are in research studies. These bad things are called "risks." The risks of being in this study might be embarrassment due to singing alone or listening to your own voice. Not all of these things may happen to you. None of them may happen. Things may happen that I don't know about yet. If they do, I will make sure that you get help to deal with anything bad that might happen.

Are there any good things that might happen during the research study?

Sometimes good things happen to people who are in research studies. These good things are called "benefits." The benefits of being in this study might be you might sing more

in-tune, or listen better to your own voice. I don't know for sure if you will have any benefits. Hopefully there will be benefits that help all of the music teachers help their students to sing better.

Who can I ask if I have any questions?

If you have any questions about this study, you can ask your parents or guardians. Also, if you have any questions that you didn't think of now, you can ask later by calling Dylan Fixmer at (828) 335-5444 or by email at dfixmer@indiana.edu. You can also email Dr. Peter Miksza at pmiksza@indiana.edu.

What if I don't want to be in the study?

If you don't want to be in this study, you don't have to. It's up to you. If you say you want to be in it and then change your mind, that's okay too. All you have to do is tell me that you don't want to be in it anymore. No one will be mad at you or upset with you if you don't want to be in it.

If you write my name on the line below, it means that you agree to be in this research study.

| | |
|---------------------|------|
| Subject's signature | Date |
|---------------------|------|

Subject's printed name

| | |
|--------------------------------------|------|
| Signature of person obtaining assent | Date |
|--------------------------------------|------|

Name of person obtaining assent

Form date: September 19, 2016

Appendix B**Singing Voice Development Measure (SVDM: Version 2) (Rutkowski, 1990a)**

- 1 "Pre-singer" does not sing but chants the song text
- 2 "Speaking Range Singer" sustains tones and exhibits some sensitivity to pitch
but remains in the speaking voice range (usually A2 to C3)
- 3 "Uncertain Singer" waivers between speaking and singing voices, uses a
limited range when in singing voice (usually up to F3)
- 4 "Initial Range Singer" exhibits use of initial singing range (usually G3 to A3)
- 5 "Singer" exhibits use of extended range (sings beyond the register lift: 83-flat
and above)

Appendix C

Song Teaching Procedure

1. Researcher will sing "Busy Buzzy Bee" in falsetto using a puppet and have the students listen.
2. Researcher will explain that this song is a game where everyone is a tree and have the students stand like trees. The researcher will sing the song again and walk between the "trees." Upon finishing the song the teacher will tap one of the students and say "I was the bee that time, this time [tapped child] gets to be the tree."
3. Researcher will have the students play the game until every child has had a turn being the bee. The researcher will sing the song 15 times, encouraging students to mouth the words along without using a singing voice yet.
4. After the 15th time singing, the researcher will invite the students to sing the song without the researcher singing. If the students are unsuccessful the researcher will sing the song three more times while the students continue to play the game. If the students are successful the first time, the researcher will simply let them sing out the rest of the game. The researcher may mouth the words but will not sing with the students.

General notes: The researcher will always count off on an treble clef second space A with "1,2, ready sing" so that the students hear the starting pitch. The researcher will never sing with the students, nor the students sing with the researcher. If the students are not singing correctly the researcher will have the students listen and mouth the words while the researcher sings alone (this is to provide an accurate model at all times).

Appendix D**Recording Booth Pictures**

Appendix E**Data Set Sample**

| Note | Correct Pitches | P1-1 Hz | Dev (in Cents) | Adj Dev (in cents) | Adj Abs Dev |
|------|-----------------|---------|----------------|--------------------|-------------|
| A | 440 | 395.5 | -184.5909949 | -5.608479221 | 5.608479221 |
| B | 493.88 | 448.3 | -167.6352571 | 11.34725851 | 11.34725851 |
| A | 440 | 390.6 | -206.1738894 | -27.19137379 | 27.19137379 |
| F# | 369.99 | 334.5 | -174.5760811 | 4.406434533 | 4.406434533 |
| A | 440 | 387.3 | -220.8624269 | -41.87991122 | 41.87991122 |
| B | 493.88 | 449 | -164.9341224 | 14.0483932 | 14.0483932 |
| A | 440 | 395.2 | -185.9046922 | -6.922176552 | 6.922176552 |
| A | 440 | 395.6 | -184.1533172 | -5.170801542 | 5.170801542 |
| B | 493.88 | 451.7 | -154.5547599 | 24.42775577 | 24.42775577 |
| A | 440 | 394 | -191.1694729 | -12.18695724 | 12.18695724 |
| F# | 369.99 | 331.7 | -189.1287196 | -10.14620396 | 10.14620396 |
| A | 440 | 424.2 | -63.31068346 | 115.6718322 | 115.6718322 |
| B | 493.88 | 476.5 | -62.02119944 | 116.9613162 | 116.9613162 |
| A | 440 | 396 | -182.4037121 | -3.421196493 | 3.421196493 |
| A | 440 | 403 | -152.068422 | 26.91409365 | 26.91409365 |
| B | 493.88 | 441.1 | -195.6657439 | -16.68322824 | 16.68322824 |
| A | 440 | 398.5 | -171.5085594 | 7.4739562 | 7.4739562 |
| F# | 369.99 | 332 | -187.5636443 | -8.58112862 | 8.58112862 |
| A | 440 | 394.9 | -187.2193871 | -8.236871501 | 8.236871501 |
| B | 493.88 | 440 | -199.9884279 | -21.00591226 | 21.00591226 |
| A | 440 | 392.6 | -197.3320211 | -18.34950549 | 18.34950549 |
| A | 440 | 397.5 | -175.858396 | 3.124119618 | 3.124119618 |
| B | 493.88 | 441.8 | -192.920554 | -13.93803833 | 13.93803833 |
| A | 440 | 389.9 | -209.2792428 | -30.29672718 | 30.29672718 |
| G | 392 | 357.1 | -161.4306245 | 17.55189114 | 17.55189114 |
| F# | 369.99 | 332.7 | -183.9172947 | -4.934779042 | 4.934779042 |

| | | | | | |
|---------------------------|---------------------------------|-------------|--------------|--------------|-------------------------------|
| E | 329.63 | 293.1 | -203.3454981 | -24.36298249 | 24.36298249 |
| D | 293.66 | 265.3 | -175.8265094 | 3.156006225 | 3.156006225 |
| Average A (hz) | 440 | 396.7833333 | | | Av Adj Dev (cents) |
| | Av Pitch Dev (cents) | 178.9825156 | | | 21.57140466 |